

Claims

What is claimed is:

1. A computer-based method of mining one or more patterns in an input data set of items, the method comprising the steps of:

5 identifying one or more sets of items in the input data set as one or more patterns based on whether the one or more sets respectively satisfy a dependency test, the dependency test being satisfied when each of the items in a set of items is dependent upon each other item with a prescribed significance level; and

10 outputting the one or more identified patterns based on results of the dependency tests.

2. The method of claim 1, wherein the dependency test employs a normal approximation test when an occurrence count of the items of a set is above a threshold value, and a Poisson approximation test otherwise.

15 3. The method of claim 1, wherein a minimum support threshold value associated with the dependency test increases as the frequency of items in a set increases, when a probability that the set is in the input data set is less than a predetermined percentage.

4. The method of claim 3, wherein the predetermined percentage is approximately fifty percent.

20 5. The method of claim 1, wherein a minimum support threshold value associated with the dependency test decreases as the size of an item set increases.

6. The method of claim 1, wherein the input data set comprises transaction data.

7. The method of claim 1, wherein the input data set comprises event data.

8. A computer-based method of mining one or more patterns in an input data set of items, the method comprising the steps of:

obtaining an input data set of items;

searching the input data set of items to identify one or more sets of items in the input data set as one or more patterns based on whether the one or more sets respectively satisfy a dependency test, the dependency test being satisfied when each of the items in a set of items is dependent upon each other item with a prescribed significance level; and

outputting the one or more identified patterns based on results of the dependency tests.

9. The method of claim 8, further comprising, prior to the searching step, the step of normalizing the input data set.

10. The method of claim 9, wherein the input data set comprises event data and the normalizing step comprises transforming at least a portion of the event data into event classes such that the event data is non-application-dependent.

11. The method of claim 10, wherein the event data transformation step further comprises the step of mapping two or more attributes associated with an event into an event class.

12. The method of claim 11, wherein the mapping step is performed in accordance with a lookup table.

13. The method of claim 10, wherein the event data is in a tabular form with a first number of columns before the transformation step and in a tabular form with a

second number of columns after the transformation step, the second number of columns being less than the first number of columns.

14. The method of claim 8, wherein the outputting step further comprises converting the one or more identified patterns into a human readable format.

5 15. The method of claim 8, wherein the searching step further comprises the step of performing a level-wise scan based on a set length to determine candidate sets of items in the input data set that satisfy the dependency test.

16. The method of claim 15, further comprising the step of pruning candidate sets.

10 17. Apparatus for mining one or more patterns in an input data set of items, the apparatus comprising:

15 at least one processor operative to: (i) identify one or more sets of items in the input data set as one or more patterns based on whether the one or more sets respectively satisfy a dependency test, the dependency test being satisfied when each of the items in a set of items is dependent upon each other item with a prescribed significance level; and (ii) output the one or more identified patterns based on results of the dependency tests; and

a memory, coupled to the at least one processor, which stores at least one of the input data set and the one or more identified patterns.

20 18. The apparatus of claim 17, wherein the dependency test employs a normal approximation test when an occurrence count of the items of a set is above a threshold value, and a Poisson approximation test otherwise.

19. The apparatus of claim 17, wherein a minimum support threshold value associated with the dependency test increases as the frequency of items in a set increases, when a probability that the set is in the input data set is less than a predetermined percentage.

5 20. The apparatus of claim 19, wherein the predetermined percentage is approximately fifty percent.

21. The apparatus of claim 17, wherein a minimum support threshold value associated with the dependency test decreases as the size of an item set increases.

10 22. The apparatus of claim 17, wherein the input data set comprises transaction data.

23. The apparatus of claim 17, wherein the input data set comprises event data.

24. Apparatus for mining one or more patterns in an input data set of items, the apparatus comprising:

15 at least one processor operative to: (i) obtain an input data set of items; (ii) search the input data set of items to identify one or more sets of items in the input data set as one or more patterns based on whether the one or more sets respectively satisfy a dependency test, the dependency test being satisfied when each of the items in a set of items is dependent upon each other item with a prescribed significance level; and (iii) output the one or more identified patterns based on results of the dependency tests; and

20 a memory, coupled to the at least one processor, which stores at least one of the input data set and the one or more identified patterns.

25. The apparatus of claim 24, further comprising, prior to the searching operation, the operation of normalizing the input data set.

26. The apparatus of claim 25, wherein the input data set comprises event data and the normalizing operation comprises transforming at least a portion of the event data into event classes such that the event data is non-application-dependent.

27. The apparatus of claim 26, wherein the event data transformation operation further comprises the operation of mapping two or more attributes associated with an event into an event class.

28. The apparatus of claim 27, wherein the mapping operation is performed in accordance with a lookup table.

29. The apparatus of claim 26, wherein the event data is in a tabular form with a first number of columns before the transformation operation and in a tabular form with a second number of columns after the transformation operation, the second number of columns being less than the first number of columns.

30. The apparatus of claim 24, wherein the outputting operation further comprises converting the one or more identified patterns into a human readable format.

31. The apparatus of claim 24, wherein the searching operation further comprises the operation of performing a level-wise scan based on a set length to determine candidate sets of items in the input data set that satisfy the dependency test.

32. The apparatus of claim 31, further comprising the operation of pruning candidate sets.

33. An article of manufacture for mining one or more patterns in an input data set of items, the article comprising a machine readable medium containing one or more programs which when executed implement the steps of:

identifying one or more sets of items in the input data set as one or more patterns based on whether the one or more sets respectively satisfy a dependency test, the dependency test being satisfied when each of the items in a set of items is dependent upon each other item with a prescribed significance level; and

outputting the one or more identified patterns based on results of the dependency tests.

34. An article of manufacture for mining one or more patterns in an input data set of items, the article comprising a machine readable medium containing one or more programs which when executed implement the steps of:

obtaining an input data set of items;

searching the input data set of items to identify one or more sets of items in the input data set as one or more patterns based on whether the one or more sets respectively satisfy a dependency test, the dependency test being satisfied when each of the items in a set of items is dependent upon each other item with a prescribed significance level; and

outputting the one or more identified patterns based on results of the dependency tests.